On second-order Taylor expansion of critical values

Bütikofer, Stephan ; Klatte, Diethard ; Kummer, Bernd

Abstract: Studying a critical value function in parametric nonlinear programming, we recall conditions guaranteeing that is a $C^{1,1}$ function and derive second order Taylor expansion formulas including second-order terms in the form of certain generalized derivatives of D. Several specializations and applications are discussed. These results are understood as supplements to the well–developed theory of first- and second-order directional differentiability of the optimal value function in parametric optimization.

Originally published at:
Bütikofer, Stephan; Klatte, Diethard; Kummer, Bernd (2010). On second-order Taylor expansion of critical values. Kybernetika, 46(3):472-487.

KYBERNETIKA — VOLUME *46* (2010), NUMBER 3, PAGES 472–487

ON SECOND–ORDER TAYLOR EXPANSION OF CRITICAL VALUES

Stephan Bütikofer, Diethard Klatte and Bernd Kummer

Studying a critical value function φ in parametric nonlinear programming, we recall conditions guaranteeing that φ is a $C^{1,1}$ function and derive second order Taylor expansion formulas including second-order terms in the form of certain generalized derivatives of $D\varphi$. Several specializations and applications are discussed. These results are understood as supplements to the well–developed theory of first- and second-order directional differentiability of the optimal value function in parametric optimization.

Keywords: Taylor expansion, parametric programs, critical value function, generalized derivatives, envelope theorems, Lipschitz stability, $C^{1,1}$ optimization

Classification: 49J52, 49K40, 90C31, 65K10

1. INTRODUCTION

In this paper, we study a parametric nonlinear program of the type

$$P(t),\ t \in T: \qquad \min_x f(x,t) \quad \text{s.t.} \quad g(x,t) \le 0, \tag{1.1}$$

where T is an open neighborhood of some $t^0 \in \mathbb{R}^l$, and f and g map $\mathbb{R}^{n+l}$ to $\mathbb{R}$ and $\mathbb{R}^m$, respectively. If nothing else is said, we take $t^0 = 0$, and suppose that f and $g = (g_1, \dots, g_m)$ are C^1 functions with locally Lipschitz derivatives (briefly $f, g \in C^{1,1}$). For given t, let $M(t)$ be the *feasible set of* $P(t)$ and denote by $S(t)$ the set of all *critical points of* $P(t)$,

$$S(t) := \{(x,y) \in \mathbb{R}^{n+m} \mid F(x,y,t) = 0\}, \tag{1.2}$$

where $F(x,y,t) = 0$ describes Kojima's [20] equivalent reformulation of the Karush-Kuhn-Tucker conditions for $P(t)$, and $F = (F_1, F_2)$ is defined by

$$\begin{aligned} F_1(x,y,t) &= D_x f(x,t) + D_x g(x,t)^{\mathsf T} y^+ \\ F_2(x,y,t) &= g(x,t) - y^- \ , \end{aligned} \tag{1.3}$$

where y^+, y^- are the vectors with components $y_i^+ = \max\{y_i, 0\}$, $y_i^- = \min\{y_i, 0\}$, and Df (resp. $D_x f$, $D_t f$) denote the derivative (resp. the partial derivatives) of f, similarly for g. We define the *Lagrange function* related to (1.1) by

$$L(x,y,t) := f(x,t) + g(x,t)^{\mathsf T} y,\ (x,y) \in \mathbb{R}^{n+m}. \tag{1.4}$$

To get a compact and brief description of our results, we do not consider additional equality constraints of $h_j(x,t) = 0$, $j = 1, \dots, \kappa$. As far as we work with Kojima's function F, it can be easily and directly seen by the assumptions and proofs below that the equalities play the same role as inequalities with positive multiplier components y_i^0 of some "nperturbed" critical point $s^0 = (x^0, y^0)$ to $P(0)$.

Since we are interested in local properties of critical values, we will restrict ourselves to an open neighborhood $\mathcal{O}$ of s^0. Given any selection $s(t) = (x(t), y(t)) \in S(t) \cap \mathcal{O}$, $t \in T$, we say that $\varphi(x(t), t)$, $t \in T$, is the associated *critical value function.* If $x(t)$ is even a (global) minimizer of $P(t)$, then $\varphi(\cdot)$ coincides with the *optimal value function* of problem (1.1), provided a constraint qualification is satisfied.

The present paper is devoted to differential stability of optimization problems, but it has a particular focus: we study conditions guaranteeing that a critical value function φ belongs (locally) to the class $C^{1,1}$, thereby deriving formulas for the Taylor expansion of the critical value function φ with second–order terms of the form $TD\varphi$ (Thibault derivative) and $CD\varphi$ (contingent derivative). This analysis is done under the assumption of *strong regularity* of the critical point map in the sense of Robinson [29]. Our approach is essentially based on the use of implicit multifunction theory (which is common in the literature), but without assuming that the data are C^2 functions or that the critical points under consideration represent local or global minimizers.

The studies presented here are understood as supplements to the well–developed theory of first- and second-order directional differentiability of the optimal value function, for which we refer exemplarily to the books or surveys [1, 5, 6, 8, 10, 11, 27, 30] with many references to the field. Some results given below are known (cf. e. g. [17, 23]) or could be derived from the literature just mentioned. On the other hand, since the differential stability under investigation is important both in theory and applications, we think that it is of value to have a self-contained presentation. For example, results of this type are of interest in convergence studies of nonsmooth Newton methods for solving critical point systems and in optimality and stability analysis in certain bi-level problems.

Some basic preliminaries from variational analysis are compiled in Section 2. Formulas for the Thibault and contingent derivatives of the gradient of a critical value function, both for smooth and canonical perturbations, are derived in § 3 and discussed in general and special settings. Section 4 is devoted to an application in bi-level optimization, which motivates and illustrates in detail our analysis.

2. PRELIMINARIES

2.1. Generalized derivatives of locally Lipschitz functions

Here we compile some facts on generalized (directional) derivatives of locally Lipschitz functions, for details see e. g. [17, Chapters 5-6].

Throughout $\|\cdot\|$ and B denote any norm resp. the related closed unit ball in some finite-dimensional linear space. Let ψ be a given locally Lipschitz function from $\mathbb{R}^l$ to $\mathbb{R}^q$, briefly $\psi \in C^{0,1}(\mathbb{R}^l, \mathbb{R}^q)$ or simply $\psi \in C^{0,1}$, and let $t^0, \tau \in \mathbb{R}^l$. For any neighborhood $\mathcal{N}$ of t^0, we put $\operatorname{Lip}(\psi, \mathcal{N}) := \inf\{L \mid \|\psi(t') - \psi(t)\| \le L\|t' - t\| \ \forall\, t, t' \in$

$\mathcal{N}\}$. The set $T\psi(t^0)(\tau)$ consisting of all limits of the form $\theta^{-1}[\psi(t+\theta\tau)-\psi(t)]$ attainable with certain sequences $\theta \downarrow 0$ and $t \to t^0$ is called the *Thibault derivative* [34] of ψ at t in direction τ. Similarly, the *contingent derivative* $C\psi(t^0)(\tau)$ consists of all limits of of $\theta^{-1}[\psi(t^0+\theta\tau)-\psi(t^0)]$ attainable with certain sequences $\theta \downarrow 0$. Generalizations of these notions to continuous functions and multivalued mappings play also an important rule in variational analysis, see e. g. [17, 30], but we do not use them here.

In this setting, $T\psi(t^0)(\tau)$ and $C\psi(t^0)(\tau)$ are nonempty, compact and connected sets, and one has $\operatorname{conv} T\psi(t^0)(\tau) = \partial\psi(t^0)\tau$, where $\partial\psi(t^0)$ is Clarke's [4] generalized Jacobian of ψ at t^0 and $\operatorname{conv} X$ denotes the convex hull of X. Hence, if ψ is real-valued then $T\psi(t^0)(\tau) = \partial\psi(t^0)\tau$. If $C\psi(t^0)(\tau)$ is a singleton, it coincides with the standard directional derivative $\psi'(t^0;\tau)$. For $X \subset \mathbb{R}^l$, we write $C\psi(t)(X) := \{C\psi(t)\sigma \mid \sigma \in X\}$, similarly for $T\psi$. For a small perturbation $f \in C^{0,1}$ with $f(t^0) = 0$, one has (cf. [17, Lemma 6.3])

$$\begin{array}{ll} (i) & \text{if } \operatorname{Lip}(f, t^0+\rho B) = O(\rho), \text{ then } T(\psi+f)(t^0)(\tau) = T\psi(t^0)(\tau), \\ (ii) & \text{if } f(t) = o(t-t^0), \text{ then } C(\psi+f)(t^0)(\tau) = C\psi(t^0)(\tau), \end{array} \tag{2.1}$$

where as usual $O(\cdot)$ means $O(\rho) \to 0$ as $\rho \to 0$ and $O(0) = 0$, while o-type functions $o(\cdot)$ satisfy $o(u)/\|u\| \to 0$ as $\|u\| \to 0$ and $o(0) = 0$.

Given $\psi \in C^{0,1}(\mathbb{R}^l, \mathbb{R}^q)$ and $\eta \in C^{0,1}(\mathbb{R}^q, \mathbb{R}^d)$, chain rules for $h(t) := \eta(\psi(t))$ and $s = \psi(t)$ are generally generally only valid as inclusions $Ch(t)(\tau) \subset C\eta(s)\,(C\psi(t)(\tau))$, similarly for $Th(t)(\tau)$. Under additional assumptions, equality holds true,

$$\begin{array}{lll} \psi \text{ or } \eta \text{ directionally differentiable} & \Rightarrow & Ch(t)(\tau) = C\eta(s)\,(C\psi(t)(\tau)) \\ \eta \in C^1 & \Rightarrow & Th(t)(\tau) = D\eta(s)\,(T\psi(t)(\tau)), \end{array} \tag{2.2}$$

see e. g. §6.4.1 in [17]. For $y, v \in \mathbb{R}^m$, define

$$\begin{array}{rcl} I(y,v) & := & \{i \in \{1,\dots,m\} \mid y_i > 0 \text{ or } y_i = 0 \le v_i\}, \\ \mathcal{R}_C(y,v) & := & \{r \in \{0,1\}^m \mid r_i = 1 \text{ if } i \in I(y,v),\ r_i = 0 \text{ else}\}, \\ \mathcal{R}_T(y) & := & \{r \in [0,1]^m \mid r_i = 1 \text{ if } y_i > 0,\ \ r_i = 0 \text{ if } y_i < 0\}. \end{array}$$

Obviously, $\mathcal{R}_C(y,v)$ is a singleton, say $r(y,v)$, and for the function $h(y) = y^+$,

$$\begin{array}{l} Ch(y)(v) = h'(y;v) = \{(r_1(y,v)\,v_1\,,\dots,\,r_m(y,v)\,v_m)\} = R_C(y,v)\circ v, \\ Th(y)(v) = \partial h(y)\,v = \{(r_1\,v_1,\dots,r_m\,v_m) \mid r \in \mathcal{R}_T(y)\} = R_T(y)\circ v, \end{array} \tag{2.3}$$

where $r\circ v = (r_1, v_1, \dots, r_m v_m)$ denotes the *Schur(-Hadamard) product* of $r, v \in \mathbb{R}^m$.

2.2. Taylor expansion and growth conditions for $C^{1,1}$ functions

Consider a function $\psi \in C^{1,1}(\mathbb{R}^n, \mathbb{R})$ and points $x^0, x \in \mathbb{R}^n$ as well as the Euclidean inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$.

Then the Thibault derivative allows the following second-order Taylor expansion at the point x, see [22, Thm. 3] or [17, Thm. 6.20]: There exist some $\theta \in (0,1)$ and some $q \in T[D\psi](x+\theta u)(u)$ such that

$$\psi(x+u) - \psi(x) - D\psi(x)u = \frac{1}{2}\langle u, q\rangle.$$

This is completely analogous to the related statement when using Clarke's generalized Jacobian $\partial[D\psi]$ instead of $T[D\psi]$, we refer to [12].

As a consequence of this Taylor formula one obtains a uniform quadratic estimate: Let c be a constant such that

$$c < \inf\{\langle w, q\rangle \mid q \in T[D\psi](x)(w)\} \quad \forall\, w:\ \|w\| = 1.$$

Then there exist some neighborhood $\mathcal{N}$ of x such that

$$\psi(\xi') - \psi(\xi) - D\psi(\xi)(\xi' - \xi) \geq \frac{1}{2}c\|\xi' - \xi\|^2 \quad \forall\, \xi,\, \xi' \in \mathcal{N},$$

for a proof we refer to [17, Cor. 6.21]. This can be interpreted in the case $c > 0$ as a growth condition *around the given point* x of interest.

In contrast, a Taylor expansion in terms of the contingent derivative is not possible, in general. However, again some quadratic estimate holds true, which can be now interpreted for $c > 0$ as a quadratic growth condition *at the given point* x of interest: Let c be a constant such that

$$c < \inf\{\langle w, q\rangle \mid q \in C[D\psi](x)(w)\} \quad \forall\, w:\ \|w\| = 1.$$

Then there exist some $\delta > 0$ such that

$$\psi(x + u) - \psi(x) - D\psi(x)u \geq \frac{1}{2}c\|u\|^2 \quad \forall\, u:\ \|u\| \leq \delta,$$

for a proof we refer to [17, Thm. 6.23].

2.3. Stability of critical points

For computing derivatives and a generalized Taylor expansion of critical values, one needs descriptions of generalized derivatives of the Kojima function, and one has to handle local Lipschitz stability properties of the critical point map S of the parametric problem (1.1) and of the critical point map S_0 of the *canonically perturbed* pendant to (1.1),

$$P_0(p),\ p = (a, b) \in \mathbb{R}^{n+m}: \qquad \min_x\ f(x, 0) - \langle a, x\rangle \quad \text{s.t.} \quad g(x, 0) \leq b, \tag{2.4}$$

where $f, g \in C^{1,1}$ is assumed. Define the Kojima function resp. the critical point map for (1.2) by

$$F_0(x, y) := F(x, y, 0) \quad \text{resp.} \quad S_0(p) := \{(x, y) \mid F_0(x, y) = p\}. \tag{2.5}$$

Then the Thibault derivative TF_0 of F_0 at $s^0 = (x^0, y^0) \in \mathbb{R}^{n+m}$ in direction (u, v) can be described by

$$TF_0(s^0)(u, v) = \left\{ \left(\begin{array}{c} q(u) + \sum_{i=1}^m r_i v_i D_x g_i(x^0, 0) \\ D_x g_i(x^0, 0)u\ -\ (1 - r_i)v_i\ (\forall\, i) \end{array} \right) \,\middle|\, \begin{array}{l} q(u) \in Q_T(u) \\ r \in \mathcal{R}_T(y^0) \end{array} \right\}, \tag{2.6}$$

where $Q_T(u) := T_x[D_xL](x^0, y^{0+}, 0)(u)$, while the contingent derivative $CF_0(s^0)(u,v)$ can be represented in the form, with $Q_C(u) := C_x[D_xL](x^0, y^{0+}, 0)(u)$,

$$CF_0(s^0)(u,v) = \left\{ \left(\begin{array}{ll} q(u) + \sum_{i\in I(y^0,v)} v_i D_x g_i(x^0,0) & \\ D_x g_i(x^0,0)u & (i \in I(y^0,v)) \\ D_x g_i(x^0,0)u \; - \; v_i & (i \notin I(y^0,v)) \end{array} \right) \middle| \; q(u) \in Q_C(u) \right\}. \tag{2.7}$$

For proving (2.6) and (2.7) see [16] or [17, Thm. 7.6].

Given $s^0 = (x^0, y^0) \in S(0)$, we call the problem (1.1) *strongly Lipschitz stable at* $(s^0, 0)$ if there are neighborhoods $\mathcal{O}$, $\mathcal{N}$ of s^0 and 0, respectively, and some function $s(\cdot) = (x(\cdot), y(\cdot))$ such that

$$S(t) \cap \mathcal{O} = \{s(t)\},\ t \in \mathcal{N},\ \ s(0) = s^0,\ \text{ and } s(\cdot) \text{ is Lipschitz on } \mathcal{N}, \tag{2.8}$$

and we say that F_0 *is strongly regular at* $(s^0, \underline{0})$ (or, synonymously, s^0 *is a strongly regular critical point of* $P(0)$) if there are neighborhoods U, V of s^0 and $\underline{0} = (0,0) \in \mathbb{R}^{n+m}$ such that $F_0^{-1}(\cdot) \cap \mathcal{U}$ is single-valued and Lipschitz on V. A relation between both notions is given in [29], see also [17, Cor. 8.4]:

Proposition 2.1. If $D_tF(\cdot,\cdot)$ exist and is Lipschitzian near $(s^0, 0)$, then (1.1) is strongly Lipschitz stable at $(s^0, 0)$ if F_0 is strongly regular at $(s^0, \underline{0})$.

For characterizations of strong regularity via injectivity of TF_0 see e.g. [17, Theorem 8.2]. Derivative-free characterizations of strong regularity and other stability properties, are e.g. given in [17, 18, 26].

3. GENERALIZED SECOND–ORDER DERIVATIVES OF CRITICAL VALUES

3.1. Differentiability of the critical value function

Now we recall a basic result on Fréchet differentiability of critical value functions. We again consider the parametric program (1.1) and suppose that

$$s^0 = (x^0, y^0) \in S(0) \cap \mathcal{O},\ \ \mathcal{O} \text{ neighborhood of } s^0.$$

The following theorem was first given by Jongen et al. in 1986, for completeness we adapt the short proof. Conditions ensuring property (3.1) for optimal values $\varphi(t)$, appear in the economic literature as *envelope theorems*, cf., e.g., [33, 35] for related applications.

Theorem 3.1. (Jongen, Möbert and Tammer [14, Lemma 2.1]) Let $f, g \in C^1$, and suppose that $(x(t), y(t)) \in S(t) \cap \mathcal{O}$, $t \in T$, is a function such that $(x(0), y(0)) = (x^0, y^0)$, $y(\cdot)$ being continuous at $t = 0$, and $x(\cdot)$ being upper Lipschitz at $t = 0$ with modulus $c > 0$, i.e., $\|x(t) - x^0\| \le c\|t\|$ for t near 0. Then the associated critical value function $\varphi = f(x(\cdot), \cdot)$ is Fréchet differentiable at $t = 0$, and its gradient $D\varphi(0)$ has the form

$$D\varphi(0) = D_tL(x^0, y^{0+}, 0). \tag{3.1}$$

P r o o f. By assumption, $g(x^0,0)^\mathsf{T} y^{0+} = (y^{0-})^\mathsf{T} y^{0+} = 0$. Since $y(\cdot)$ is continuous at $t=0$, we have for t near 0, that $y_j^0 = y_j(0) > 0$ implies $y_j(t) > 0$ (hence $g_j(x(t),t) = y_j(t)^- = 0$), and therefore

$$(g(x(t),t) - g(x^0,0))^\mathsf{T} y^{0+} = 0. \tag{3.2}$$

Since $L(\cdot, y^{0+}, \cdot) \in C^1$, one has

$$L(x,y^{0+},t) - L(x^0,y^{0+},0) - D_{(x,t)}L(x^0,y^{0+},0)\begin{pmatrix} x-x^0 \\ t-0 \end{pmatrix} = o(x-x^0, t-0). \tag{3.3}$$

Let $\varepsilon > 0$ be arbitrary. Then we have for t sufficiently close to 0, by using (3.2) and $D_x L(x^0,y^{0+},0) = 0$, as well as (3.3) and the upper Lipschitz property of $x(\cdot)$,

$$\begin{aligned}
&|\,\varphi(t) - \varphi(0) - D_t L(x^0,y^{0+},0)(t-0)\,| \\
&= |\, f(x(t),t) - f(x^0,0) + (g(x(t),t) - g(x^0,0))^\mathsf{T} y^{0+} \\
&\qquad\qquad - D_t L(x^0,y^{0+},0)(t-0) - D_x L(x^0,y^{0+},0)(x(t)-x^0)\,| \\
&= |\, L(x(t),y^{0+},t) - L(x^0,y^{0+},0) - D_{(x,t)}L(x^0,y^{0+},0)\tbinom{x(t)-x^0}{t-0}\,| \\
&= |o(x(t)-x^0, t-0)| \;\le\; \varepsilon\,\|(x(t)-x^0, t-0)\|_1 \;\le\; \varepsilon(c+1)\|t\|_1,
\end{aligned}$$

where $\|\cdot\|_1$ denotes the sum norm both for $\mathbb{R}^{n+l}$ and $\mathbb{R}^l$. Since ε was arbitrary, this shows that $\varphi(t) - \varphi(0) - D_t L(x^0,y^{0+},0)t = o(t)$ for some (new) o-function. This completes the proof. ☐

In the following remarks we use for selected CQs (constraint qualifications) the abbreviations *LICQ* (Linear Independence CQ), *MFCQ* (Mangasarian–Fromovitz CQ) or *strict MFCQ*, their definitions are standard, cf. [1, 7, 17].

Remark 3.2. Note that the existence of a continuous selection $y(\cdot)$ of the multiplier mapping is essential, and in general MFCQ is not enough to guarantee the assertion of Theorem 3.1. Consider the parametric linear program for $t \in T = (-1,1)$,

$$\min\; x_1 + x_2 + tx_1 \quad \text{s.t.} \quad -x_1 - x_2 \le t,\; x_1, x_2 \ge 0.$$

Its optimal value function φ is not differentiable at $t=0$. One easily checks: For $t \in T$, (i) $\varphi(t) \equiv 0$ if $t \ge 0$, but $\varphi(t) = -t - t^2$ if $t < 0$, (ii) for $t \ne 0$, there is a unique primal solution $x(t) = (0,0)$ if $t > 0$ and $x(t) = (-t,0)$ if $t < 0$, and the unique dual solutions $y(t)$ satisfy $y(t) \to (0,1,1)$ as $t \downarrow 0$, but $y(t) \to (1,0,0)$ as $t \uparrow 0$, (iii) MFCQ is satisfied everywhere, and (iv) at $t=0$, $x(\cdot)$ is upper Lipschitz, but $y(\cdot)$ is discontinuous.

Remark 3.3. If the critical point map $S(t)$ is *locally nonempty-valued and upper Lipschitz at* $(0,(x^0,y^0))$, i. e., if for some constant $L > 0$ and some neighborhoods $\mathcal{O}$ of $s^0 = (x^0,y^0)$ and $\mathcal{N}$ of $t=0$,

$$\emptyset \ne S(t) \cap \mathcal{O} \subset s^0 + L\|t\|B \quad (\forall\, t \in \mathcal{N}), \tag{3.4}$$

then, obviously, any $(x(t), y(t)) \in S(t) \cap \mathcal{O}$ satisfies the assumptions of Theorem 3.1. In order to guarantee (3.4) in the case $f, g \in C^2$ it is e. g. sufficient that x^0 is a strict local minimizer of $P(0)$ satisfying both the strict MFCQ (hence there is a unique associated multiplier y^0) and a standard second-order sufficient optimality condition, cf. [17, Cor. 8.16, Thm. 8.36].

Remark 3.4. For the case that $x(t)$ are global minimizers of $P(t)$ under $f, g \in C^1$, there is a well-known envelope theorem first given by Gauvin and Dubeau [9], see also slight modifications of it in [4, §6.5] (for special perturbations) and [33, Thm. 3.8.4]:

Theorem 3.1'. If x^0 is a unique global minimizer of $P(0)$, if $M(\cdot)$ is uniformly compact near $t = 0$ and if LICQ holds at x^0 w.r. to $M(0)$, then the optimal value function $\varphi(t) = \inf_x\{f(x,t) \,|\, g(x,t) \le 0\}$ is Fréchet differentiable at $t = 0$ and (3.1) holds true.

Note that, under the assumptions of Theorem 3.1', the multiplier y^0 associated with x^0 is unique, and there are neighborhoods $\mathcal{N}, \mathcal{O}$ of 0 and $s^0 = (x^0, y^0)$, respectively, and a function $t \in \mathcal{N} \mapsto s(t) = (x(t), y(t)) \in S(t) \cap \mathcal{O}$ being continuous at $t = 0$, where $s(0) = s^0$ and $x(t)$ is a global minimizer of $P(t)$ for $t \in \mathcal{N}$. Indeed, to obtain existence of $x(\cdot)$ with these properties, apply for example [17, Thm. 1.16] together with compactness of $M(t)$ and $\operatorname{argmin}_{x \in M(0)} f(x, 0) = \{x^0\}$; further, by persistence of LICQ for t near 0, the unique multiplier $y(t)$ associated with $x(t)$ is also continuous (cf. e. g. [19, Lemma 2.1]).

Formally, Theorem 3.1' also allows that (3.1) holds though $x(t)$ is not upper Lipschitz at $t = 0$ (as required in Theorem 3.1), consider the example

$$\min_x x^4 - tx \quad \text{s.t.} \quad -1 \le x \le 1 \qquad (t \in \mathbb{R}).$$

However, in the "typical situation" that $f, g \in C^{1,1}$ and x^0 is a global minimizer which satisfies LICQ together with a 1st- or 2nd-order growth condition for $P(0)$ (i. e., $f(x, 0) \ge f(x^0, 0) + \varrho\|x - x^0\|^q$ for all $x \in M(0)$ near x^0, with $q = 1$ or $= 2$ and some $\varrho > 0$), a global minimizer $x(t)$ of $P(t)$ becomes automatically upper Lipschitz at $t = 0$, see [15, Thm. 3.3] or [1, Thm. 4.81]. The persistence of LICQ implies again continuity of the associated unique multiplier $y(\cdot)$ at $t = 0$, hence in our "typical situation" Theorem 3.1' is a special case of Theorem 3.1.

Under stronger assumptions, we have the following obvious (and known) consequence of Theorem 3.1, where $y^+(t) := (y(t))^+$:

Corollary 3.5. Let $f, g \in C^{1,1}$, and let the problem (1.1) be strongly Lipschitz stable at $(0, s^0)$. Then for some open neighborhood V of 0 and some function $s(\cdot) = (x(\cdot), y(\cdot))$ according to (2.8), the critical value function $\varphi = f(x(\cdot), \cdot)$ fulfils

$$D\varphi(t) = D_t L(x(t), y^+(t), t), \; t \in V, \tag{3.5}$$

and hence, $\varphi \in C^{1,1}$ on V.

The proof is immediate: $s(\cdot)$ satisfies the assumptions of Theorem 3.1 not only at $t = 0$ but also at each t sufficiently close to 0, and D_tL, $s(\cdot)$ are locally Lipschitz. To avoid misunderstandings note that $D_tL(x(t), y(t), t)$ in (3.5) means the partial derivative of L w.r. to the third variable vector at the triple $(x(t), y(t), t)$, $t \in V$.

3.2. Formulas under smooth parametrization

We consider the parametric problem $P(t)$ defined in (1.1) with Kojima function F (1.3), where t varies in a neighborhood T of $0 \in \mathbb{R}^l$. Throughout we assume

$$\begin{array}{l} f, g \in C^{1,1},\ s^0 = (x^0, y^0) \text{ is a given critical point of } P(0), \\ \text{and } D_tf(\cdot,\cdot),\ D_tg(\cdot,\cdot) \text{ belong to } C^{1,1} \text{ near } (x^0, 0). \end{array} \tag{3.6}$$

Hence, $D_tF(\cdot,\cdot)$ exists and is locally Lipschitz; this allow us to apply a suitable implicit function theorem under strong regularity. Suppose now the assumptions of Corollary 3.5 are satisfied with some selection $s(t) \in S(t)$ such that $s(0) = s^0$. Hence, for t in some neighborhood V of 0,

$$D\varphi(t) = D_tL(x(t), y^+(t), t) \tag{3.7}$$

and we obtain a chain and partial differentiation rule for the composite function

$$t \mapsto G(t) := (x(t), y^+(t), id(t)) \mapsto H(G(t)),\ \text{where } H(x, y, t) := D_tL(x, y, t). \tag{3.8}$$

By assumption (3.6), D_tL is C^1 near $z^0 := (x^0, y^{0+}, 0)$, and so due to (2.2),

$$\begin{array}{l} TD\varphi(0)(\tau) = D^2_{xt}L(z^0)\,Tx(0)(\tau) \ +\ D_tg(x^0, 0)\,Ty^+(0)(\tau) \ +\ D^2_{tt}L(z^0)\tau, \\ CD\varphi(0)(\tau) = D^2_{xt}L(z^0)\,Cx(0)(\tau) \ +\ D_tg(x^0, 0)\,Cy^+(0)(\tau) \ +\ D^2_{tt}L(z^0)\tau. \end{array} \tag{3.9}$$

We assumed (3.6) to arrive at comparable (and computable) formulas for $CD\varphi$ and $TD\varphi$. Note that with respect to the contingent derivative, directional differentiability of D_tL or of $s(\cdot)$ would suffice to guarantee a formula similar to (3.9) when replacing $D^2_{xt}L$ and $D^2_{tt}L$ by C_xD_tL and C_tD_tL, respectively.

Theorem 3.6. Suppose (3.6), let $s^0 = (x^0, y^0)$ be a strongly regular critical point of $P(0)$. Then for each t in a neighborhood V of 0 there is a locally unique critical point $s(t) = (x(t), y(t))$ of $P(t)$ such that $s(\cdot)$ is locally Lipschitz on V with $s(0) = s^0$, $\varphi(\cdot) = f(x(\cdot), \cdot)$ is $C^{1,1}$ on V, and for any direction $\tau \in \mathbb{R}^l$, the formulas (3.9) hold true.

Proof. Since $s^0 = (x^0, y^0)$ is strongly regular, $p \mapsto F_0(p)$ has in some neighborhood of $(\underline{0}, s^0)$ a (single-valued) locally Lipschitz inverse $s \mapsto F_0^{-1}(s)$ and, by using (3.6), Proposition 2.1 and Corollary 3.5, (1.1) is strongly Lipschitz stable at $(s^0, 0)$, and the first part of the statement is shown. The second part was shown above. □

In the second-order derivative formulas (3.9), there appear still the unknown mappings $Tx(0)$ and $Ty^+(0)$, the same for $Cx(0)$ and $Cy^+(0)$.

To find representations by known terms, we will first show that under the assumptions of Theorem 3.6,

$$Ts(0)(\tau) = T[F_0^{-1}](\underline{0})\,(-D_tF(s^0,0)\tau), \quad \text{similarly for } Cs(0)(\tau), \tag{3.10}$$

hold for the critical point map $s(t) = (x(t), y(t))$, where $\underline{0}$ is the zero in $\mathbb{R}^{n+m}$, and F_0^{-1} being the locally single-valued and Lipschitz inverse of $F_0(\cdot) = F(\cdot, 0)$ arising from strong regularity.

Using the continuity of $D_tF(\cdot,\cdot)$, we first derive local estimates of the remainder function

$$r(s,t) := F(s,t) - F(s,0) - D_tF(s^0,0)t. \tag{3.11}$$

By the mean-value theorem, there holds

$$r(s,t) = \int_0^1 \left[D_tF(s,\theta t) - D_tF(s^0,0)\right]\, t\,\mathrm{d}\theta\,.$$

Hence,

$$\|r(s,t)\| \le O(s,t)\|t\| \quad \text{with } O(s,t) \downarrow 0 \text{ as } s \to s^0 \text{ and } t \to 0, \tag{3.12}$$

and for $\delta, \varepsilon > 0$ and $(s,t) \in (s^0 + \delta\varepsilon\, B, \delta\, B)$, we thus obtain the uniform estimates

$$\operatorname{Lip}(r(s,\cdot), \delta B) \le O(\delta) \text{ and } \sup_{t\in\delta B} \|r(s,t)\| \le o(\delta) \quad \text{if } s \in s^0 + \delta\varepsilon B. \tag{3.13}$$

According to (3.11), one has for t near 0,

$$0 = F(s(t),t) = F(s(t),0) + D_tF(s^0,0)t + r(s(t),t).$$

Using the local inverse F_0^{-1}, this is

$$s(t) = F_0^{-1}(-D_tF(s^0,0)t - r(s(t),t)).$$

Since $s = s(t)$ is locally Lipschitz, the estimates (3.13) follow and so, since F_0^{-1} belongs to $C^{0,1}$ and $O(\delta)$, $o(\delta)$ vanish in (3.13), we obtain indeed (3.10) via (2.1).

Now, (3.10) allows us to rewrite the formulas (3.9) in terms of TF_0^{-1} and CF_0^{-1}. By (2.3), and with $\mathbf{1} = (1,\dots,1) \in \mathbb{R}^n$,

$$\begin{aligned} T(x(\cdot), y^+(\cdot))(0)(\tau) &= (\mathbf{1}, \mathcal{R}_T(y^0)) \circ Ts(0)(\tau), \\ C(x(\cdot), y^+(\cdot))(0)(\tau) &= (\mathbf{1}, \mathcal{R}_C(y^0, v)) \circ Cs(0)(\tau),\ v \in \Lambda(\tau), \end{aligned}$$

where $R \circ Y$ is the set-valued Schur product $R \circ Y = \{r \circ y \in \mathbb{R}^m \mid r \in R,\ y \in Y\}$ and

$$\Lambda(\tau) := \{v \in \mathbb{R}^m \mid \exists u \in \mathbb{R}^n : (u,v) \in C[F_0^{-1}](\underline{0})\,(-D_tF(s^0,0)\tau)\}.$$

Hence, we have shown

Corollary 3.7. Under the assumptions of Theorem 3.6, the formulas (3.9) can be equivalently written for $s^0 = (x^0, y^0)$ and $z^0 = (x^0, y^{0+}, 0)$ as

$$\begin{aligned} TD\varphi(0)(\tau) &= D^2_{st}L(z^0)\ [\hat{\mathcal{R}}_T(y^0) \circ T[F_0^{-1}](\underline{0})\,(-D_tF(s^0,0)\tau)] \ +\ D^2_{tt}L(z^0)\tau, \\ CD\varphi(0)(\tau) &= D^2_{st}L(z^0)\ [\hat{\mathcal{R}}_C(y^0, \Lambda(\tau)) \circ C[F_0^{-1}](\underline{0})\,(-D_tF(s^0,0)\tau)] \ +\ D^2_{tt}L(z^0)\tau, \end{aligned} \tag{3.14}$$

where $\hat{\mathcal{R}}_T(y^0) = (\mathbf{1}, \mathcal{R}_T(y^0))$ and $\hat{\mathcal{R}}_C(y^0, \Lambda(\tau)) = \{(\mathbf{1}, \mathcal{R}_C(y^0, v)),\ v \in \Lambda(\tau)\}$.

Note that by definition, one has, with $s^0 = (x^0, y^0)$,

$$\begin{aligned} (u,v) \in TF_0^{-1}(\underline{0})(\pi) \quad &\Leftrightarrow \quad \pi \in TF_0(x^0,y^0)(u,v), \\ (u,v) \in CF_0^{-1}(\underline{0})(\pi) \quad &\Leftrightarrow \quad \pi \in CF_0(x^0,y^0)(u,v). \end{aligned} \tag{3.15}$$

Thus, the preceding corollary says that $\sigma \in TD\varphi(0)(\tau)$ if and only if the inclusion

$$-D_tF(s^0,0)\tau \in TF_0(x^0,y^0)(u,v) \tag{3.16}$$

has some solution (u,v) which satisfies

$$\sigma \in D^2_{st}L(x^0,y^{0+},0)\begin{pmatrix} u \\ r \circ v \end{pmatrix} + D^2_{tt}L(x^0,y^{0+},0)\,\tau, \quad \text{for some } r \in \mathcal{R}_T(y^0), \tag{3.17}$$

and, on the other hand, each solution (u,v) of the inclusion (3.16) defines via (3.17) an element σ of $TD\varphi(0)(\tau)$. Completely analogous formulas apply for CF_0, in particular, (3.17) has to be replaced by

$$\sigma = D^2_{xt}L(x^0,y^{0+},0)u \; + \; D_tg(x^0,0)\,[\mathcal{R}_C(y^0,v)\circ v] \; + \; D^2_{tt}L(x^0,y^{0+},0)\,\tau. \tag{3.18}$$

In the case of C^2 data, from (2.6) and (2.7) we have, again with $s^0 = (x^0,y^0)$ and $z^0 = (x^0,y^{0+},0)$,

$$TF_0(s^0)(u,v) = \left\{ \begin{pmatrix} D^2_{xx}L(z^0)\,u + \sum_{i=1}^m r_iv_iD_xg_i(x^0,0) \\ D_xg_i(x^0,0)u \; - \; (1-r_i)v_i \quad (\forall\, i) \end{pmatrix} \,\middle|\, \begin{matrix} r_i \in [0,1] & \text{if } y_i^0 = 0 \\ r_i = 1 & \text{if } y_i^0 > 0 \\ r_i = 0 & \text{if } y_i^0 < 0 \end{matrix} \right\},$$

while in computing $CD\varphi(0)(\tau)$, the inclusion (3.16) has to be replaced by the equation

$$-D_tF(s^0,0)\tau = \begin{pmatrix} D^2_{xx}L(z^0)\,u + \sum_{i\in I(y^0,v)} v_iD_xg_i(x^0,0) & \\ D_xg_i(x^0,0)u & (\text{ if } y_i^0 > 0 \text{ or } [y_i^0 = 0,\, v_i \geq 0]\,) \\ D_xg_i(x^0,0)u \; - \; v_i & (\text{ if } y_i^0 < 0 \text{ or } [y_i^0 = 0,\, v_i < 0]\,) \end{pmatrix}. \tag{3.19}$$

Finding a solution (u,v) of equation (3.19) means determining a KKT solution of some associated quadratic program, or equivalently, solving a mixed LCP. Indeed, using for i with $y_i^0 = 0$ the unique representation of v_i by a pair (α_i,β_i) via $v_i = \alpha_i + \beta_i$, $\alpha_i \geq 0 \geq \beta_i$ and $\alpha_i\beta_i = 0$ and choosing $\alpha_i = v_i$, $\beta_i = 0$ if $y_i^0 > 0$ and $\beta_i = v_i$, $\alpha_i = 0$ if $y_i^0 < 0$, (3.19) becomes the mixed LCP

$$\begin{aligned} &-D^2_{xt}L(z^0)\,\tau \; = \; D^2_{xx}L(z^0)\,u + \textstyle\sum_{i=1}^m \alpha_iD_xg_i(x^0,0) \\ &-D_tg_i(x^0,0)\,\tau = \; D_xg_i(x^0,0)\,u \; - \; \beta_i, \\ &\quad \beta_i = 0 \text{ if } y_i^0 > 0, \;\; \alpha_i = 0 \text{ if } y_i^0 < 0, \;\; [\alpha_i \geq 0 \geq \beta_i,\, \alpha_i\beta_i = 0] \text{ if } y_i^0 = 0, \end{aligned} \tag{3.20}$$

and α_i, β_i may be interpreted as multipliers resp. slack variables for a quadratic program.

Note that for a C^2 program and under the assumptions of Theorem 3.6, $s(\cdot)$ is locally a PC^1 function (i. e., a continuous selection of finitely many C^1 functions), hence $C^{0,1}$ and directionally differentiable, this is well-known, cf. e. g. [31, Thm. 4.2.2] or [17, § 8.2]. Hence, by Corollary 3.5, $D\varphi$ is also PC^1 and directionally differentiable, and so $CD\varphi$ coincides with the standard directional derivative of $D\varphi$. This means that (3.20) has a unique solution by recalling $Cs(0)(\tau) = C[F_0^{-1}](\underline{0})\,(-D_tF(s^0,0)\tau)$ from (3.10), and so we recover a classical result [13].

For general $C^{1,1}$ data the formulas are similar but with the corresponding set-valued (partial) Hessian of the Lagrangian, according to the representations (2.6) and (2.7).

3.3. Formulas under canonical perturbations

In this subsection, we consider the canonically perturbed program (2.4),

$$P_0(a,b): \quad \min\{f_0(x) - \langle a, x\rangle \,|\, g_0(x) \le b\}, \qquad (a,b) \text{ varies near } \underline{0} = (0,0) \in \mathbb{R}^{n+m},$$

at some critical point $s^0 = (x^0, y^0)$ of $P_0(0,0)$. In comparison to (1.1),

$$f_0(x) := f(x,0),\ g_0(x) := g(x,0) \qquad (f_0, g_0 \in C^{1,1}).$$

Let F_0 and S_0 be defined according to (2.5), and we suppose that F_0 is strongly regular at $(s^0, \underline{0})$, i. e., F_0^{-1} exists locally as a single-valued Lipschitz function. Define the associated critical value function by

$$\varphi_0(a,b) = f_0(x) - \langle a, x\rangle,\ (x,y) = F_0^{-1}(a,b). \tag{3.21}$$

Further recall the descriptions of the derivatives $TF_0(s^0)(u,v)$ and $CF_0(s^0)(u,v)$ in (2.6) and (2.7). Define

$$\begin{array}{l} T^-(\alpha,\beta) \text{ and } C^-(\alpha,\beta) \text{ being the sets of all } (u,v,r) \\ \text{satisfying (2.6) and (2.7), respectively.} \end{array} \tag{3.22}$$

Then, by (3.15), one has, similarly for $CF_0^{-1}(\underline{0})(\alpha,\beta)$ and $C^-(\alpha,\beta)$,

$$TF_0^{-1}(\underline{0})(\alpha,\beta) = \{(u,v)|(u,v,r) \in T^-(\alpha,\beta)\}. \tag{3.23}$$

If even $F_0 \in C^1$ near s^0, the map $(\alpha,\beta) \mapsto (u,v)$ plays the role of $DF_0(s^0)^{-1}$.

Theorem 3.8. Under strong regularity of F_0 at a critical point $s^0 = (x^0, y^0)$ of $P_0(0,0)$, the map φ_0 belongs to $C^{1,1}$, and it holds for (a,b) near $(0,0)$,

$$D\varphi_0(a,b) = -(x, y^+),\ (x,y) = F_0^{-1}(a,b). \tag{3.24}$$

Moreover,

$$\begin{aligned} TD\varphi_0(\underline{0})(\alpha,\beta) &= \{-(u\,,\, r\circ v)\,|\,(u,v,r) \in T^-(\alpha,\beta)\}, \\ CD\varphi_0(\underline{0})(\alpha,\beta) &= \{-(u\,,\, r\circ v)\,|\,(u,v,r) \in C^-(\alpha,\beta)\}, \end{aligned}$$

where $r \circ v := (r_1v_1, \ldots, r_mv_m)$ is the Schur product of r and v.

P r o o f. Writing and fixing $t = (a, b)$ (near $\underline{0}$), and putting $f(x,t) := f_0(x) - \langle a, x\rangle$ and $g(x,t) := g_0(x) - b$, all assumptions of Corollary 3.5 are satisfied at $s^0 = F_0^{-1}(a,b)$. Hence, one obtains, using also (3.21) and $L(x,y,a,b) = f_0(x) - \langle a,x\rangle + \langle y^+, g_0(x) - b\rangle$,

$$D\varphi_0(a,b) = -D_{(a,b)}\, L(x, y^+, a, b) = -(x, y^+), \ (x,y) = F_0^{-1}(a,b).$$

To show the description of $TD\varphi_0$, one has only to combine (3.23) and (3.24) and to note that the terms $r \circ v$, $r \in \mathcal{R}_T(y^0)$, just form the Thibault derivative of the function $y \mapsto y^+$, i. e., by definition of $\mathcal{R}_T(y^0)$ in §2.1,

$$\theta^{-1}[(y + \theta v)^+ - y^+] = (r_1 v_1, \dots, r_m v_m), \quad \text{(for small } \theta),$$

with $r_i \in [0,1]$ if $y_i^0 = 0$, $r_j = 1$ if $y_j^0 > 0$, $r_k = 0$ if $y_k^0 < 0$. The simple analogous proof for $CD\varphi_0$ is left to the reader. □

An alternative approach to second-order expansion of φ_0 in the special case of perturbed local minimizers under tilt perturbations (put $b \equiv 0$) can be found in [28].

Note that the above representations of $TD\varphi_0$ and $CD\varphi_0$ could be also immediately obtained from Corollary 3.7. Indeed, writing the canonical perturbations as $t = (a,b)$, $f(x,t) := f(x) - \langle a, x\rangle$, $g(x,t) := g(x) - b$, with direction $\tau = (\alpha, \beta)$, and putting $z^0 = (x^0, {y^0}^+, 0)$, one easily checks that the term $D^2_{tt}L(z^0)\tau$ vanishes, and, further, $-D_tF(s^0, 0)\tau = (\alpha, \beta)$ as well as $D^2_{st}L(z^0)(\cdot) = -\mathrm{id}\,(\cdot)$.

4. APPLICATION: SECOND–ORDER TERMS FOR A SPECIAL BILEVEL PROBLEM

As an application, we study in this section the Kojima function Φ of a special $C^{1,1}$ program in the context of bilevel problems and thereby apply the formulas of the previous sections. Consider a nonlinear program, called an *upper level problem*,

$$\min \varphi(t) \quad \text{s.t.} \quad \psi_i(t) \le 0, \ i = 1, \dots, N, \qquad (\varphi, \psi_i \in C^{1,1}), \tag{4.1}$$

where some or all of the functions φ, ψ_i are optimal or critical value functions of parametric nonlinear programs, so-called *lower level problems*. Models of the type (4.1) are classical in bi-level, semi-infinite and multi-stage optimization, we refer exemplarily to the books [5, 7, 32] and mention that there is a growing interest for this in generalized semi-infinite optimization, Nash equilibrium theory, decomposition, and other areas.

The Kojima function Φ associated with (4.1) is, similarly to (1.3),

$$\Phi_1(t,\lambda) = D\varphi(t) + D\psi(t)^{\mathsf T}\lambda^+, \quad \Phi_{2i}(t,\lambda) = D\psi_i(t) - \lambda_i^- \quad (\forall\, i),$$

where $\lambda = (\lambda_1, \dots, \lambda_N)$ is the associated multiplier vector. Then the derivatives $T\Phi$ and $C\Phi$ can be described analogously to (2.6) and (2.7), respectively, while Clarke's generalized derivative is $\partial\Phi(t)\tau = \operatorname{conv} T\Phi(t)(\tau)$. However, the corresponding descriptions are still abstract and do not utilize the bi-level structure of (4.1), and so we discuss now a special form.

For space reasons, we handle only the particular case of $\Gamma\Phi(t)(\tau) = \Phi'(t;\tau)$, and we restrict ourselves to the special setting that only φ represents a critical value function, where $\varphi(t)$ is defined via the lower level program $P(t)$ (1.1).

This is a typical situation in primal decomposition procedures: if $s^0 = (x^0, y^0)$ is a strongly regular critical point of $P(t^0)$ satisfying the assumptions (3.6) (replace there 0 by t^0) and t^0 is a stationary solution of the program (4.1) for $\varphi(t) = f(x(t), t)$, with $x(t)$ according to Theorem 3.6, then (x^0, t^0) is a stationary solution of the program $\min_{(x,t)}\{f(x,t) \mid g(x,t) \le 0,\ \psi(t) \le 0\}$; cf. [14, Thm. 2.1].

In detail, we suppose that

$$\begin{array}{l} \text{in (1.1), } t^0 \in \mathbb{R}^l \text{ and } s^0 = (x^0, y^0) \in S(t^0) \text{ are given,} \\ f, g_i \in C^2 \text{ and } D_t f,\ D_t g_i \in C^{1,1} \text{ near } (x^0, t^0), \\ s^0 \in S(t^0) \text{ is strongly regular, and by Proposition 2.1,} \\ s(t) = (x(t), y(t)) \text{ is the locally unique critical point near } (t^0, s^0), \\ \text{in (4.1) put } \varphi(t) = f(x(t), t) \text{ for } t \text{ near } t^0 \text{, and suppose } \psi_i \in C^2. \end{array} \tag{4.2}$$

Under (4.2), the assumption (3.6) (which is essential in Theorem 3.6) is automatically satisfied, and $D\varphi$ is a PC^1 function and directionally differentiable.

Define the Lagrange function of (4.1) by $\mathcal{L}(t,\lambda) := \varphi(t) + \sum_{i=1}^N \lambda_i^+\ \psi_i(t)$, hence

$$(D_t\mathcal{L})'((t,\lambda);\tau) = (D\varphi)'(t;\tau) + \sum\nolimits_{i=1}^N \lambda_i^+\ D^2\psi_i(t)\tau.$$

We then obtain, analogously to (2.7), for (t^0, λ^0), $(\tau, \mu) \in \mathbb{R}^{l+N}$,

$$\Phi'((t^0,\lambda^0);(\tau,\mu)) = \begin{pmatrix} (D\varphi)'(t^0;\tau) + \sum\limits_{i=1}^{N} (\lambda_i^0)^+ D^2\psi_i(t^0)\tau + \sum\limits_{i\in J(\lambda^0,\mu)} \mu_i D\psi_i(t^0) \\ D\psi_i(t^0)\tau \qquad\qquad (i \in J(\lambda^0,\mu)) \\ D\psi_i(t^0)\tau - \mu_i \qquad\quad (i \notin J(\lambda^0,\mu)) \end{pmatrix}, \tag{4.3}$$

where $J(\lambda^0,\mu) := \{i \,|\, \lambda_i^0 > 0\} \cup \{i \,|\ \lambda_i^0 = 0,\ \mu_i \ge 0\}$. Hence, by using the above arguments, the next theorem is immediately obtained from Corollary 3.7 and the discussion following that corollary. In particular, formula (4.4) follows by the partial differentiation rule for directional derivatives (see (3.18) and the definition of D_tL, since $s(\cdot)$ is directionally differentiable under our assumptions.

Theorem 4.1. Consider the program (4.1), suppose (4.2) and let $\tau \in \mathbb{R}^l$, $\lambda^0, \mu \in \mathbb{R}^N$ be given. Then one has on some neighborhood V of t^0 that φ is $C^{1,1}$, $D\varphi$ is directionally differentiable, and the Kojima function Φ of (4.1) is Lipschitz and directionally differentiable. Moreover, $(D\varphi)'(t^0;\tau)$ in (4.3) has the form

$$(D\varphi)'(t^0;\tau) = D^2_{xt}L(x^0, y^{0+}, t^0)u + D_t g(x^0, 0)[\,\mathcal{R}_C(y^0, v) \circ v\,] + D^2_{tt}L(x^0, y^{0+}, t^0)\tau, \tag{4.4}$$

where L is the Lagrange function (1.4) of the lower level problem (1.1), $\mathcal{R}_C(y^0, v)$ is defined according to § 2.1, $\circ$ means the Schur product, and $\sigma = (u, v)$ is the unique solution of the equation (3.20) when replacing there $(s^0, 0)$ and $(x^0, 0)$ by (s^0, t^0) and (x^0, t^0), respectively.

The assumptions of Theorem 4.1 are persistent under small perturbations, and so the formulas (4.3) and (4.4) similarly hold also for (t, λ) sufficiently close to (t^0, λ^0) and with $s(t)$ instead of s^0. This gives rise to use the following nonsmooth Newton scheme (or some inexact variant of it) for finding a zero $z^0 = (t^0, \lambda^0)$ of Φ: given an iterate z^k near z^0, compute $z^{k+1} = z^k + \zeta$ via a solution ζ of the generalized Newton equation $\Phi(z^k) + \Phi'(z^k; \zeta) = 0$, or by solving the optimization problem

$$\|\Phi(z^k) + \Phi'(z^k; \zeta)\| \to \min_{\zeta} . \tag{4.5}$$

In particular, when $\|\cdot\|$ is the Euclidean norm, one can show via the representations (4.3) and (4.4) of Φ' that the problem (4.5) becomes a program with quadratic objective function and (mixed) linear complementarity constraints, cf. [3, § 3].

When using a suitable set-valued approximation $\Gamma\Phi$ of Φ instead of Φ', the Newton step becomes $0 \in \Phi(z^k) + \Gamma\Phi(z^k; \zeta)$, and it is of interest for future research how the corresponding nonsmooth Newton schemes for standard nonlinear programs (see for such settings e. g. [2, 7, 17, 21, 24, 25]) apply to our bi-level problem.

(Received April 14, 2010)

REFERENCES

[1] J. F. Bonnans and A. Shapiro: Perturbation Analysis of Optimization Problems. Springer, New York 2000.

[2] St. Bütikofer: Globalizing a nonsmooth Newton method via nonmonotone path search. Math. Meth. Oper. Res. *68* (2008), 235–256.

[3] St. Bütikofer and D. Klatte: A nonsmooth Newton method with path search and its use in solving $C^{1,1}$ programs and semi-infinite problems. Manuscript, February 2009.

[4] F. H. Clarke: Optimization and Nonsmooth Analysis. Wiley, New York 1983.

[5] S. Dempe: Foundations of Bilevel Programming. Kluwer, Dordrecht – Boston – London 2002.

[6] V. F. Demyanov and V. N. Malozemov: Introduction to Minimax. Wiley, New York 1974.

[7] F. Facchinei and J.-S. Pang: Finite-Dimensional Variational Inequalities and Complementarity Problems, Volumes I, II. Springer, New York 2003.

[8] A. V. Fiacco: Introduction to Sensitivity and Stability Analysis. Academic Press, New York 1983.

[9] J. Gauvin and F. Dubeau: Differential properties of the marginal function in mathematical programming. Math. Program. Study *19* (1982), 101–119.

[10] J. Gauvin: Theory of Nonconvex Programming. Les Publications CRM, Montreal 1994.

[11] E. G. Golstein: Theory of Convex Programming. (Trans. Math. Monographs *36*.) American Mathematical Society, Providence 1972.

[12] J.-B. Hiriart-Urruty, J. J. Strodiot, and V. Hien Nguyen: Generalized Hessian matrix and second order optimality conditions for problems with $C^{1,1}$-data. Appl. Math. Optim. *11* (1984), 43–56.

[13] K. Jittorntrum: Solution point differentiability without strict complementarity in nonlinear programming. Math. Program. Study *21* (1984), 127–138.

[14] H. Th. Jongen, T. Möbert, and K. Tammer: On iterated minimization in nonconvex optimization. Math. Oper. Res. *11* (1986), 679–691.

[15] D. Klatte: On quantitative stability for non-isolated minima. Control and Cybernetics *23* (1994), 183–200.

[16] D. Klatte and B. Kummer: Generalized Kojima functions and Lipschitz stability of critical points. Comput. Optim. Appl. *13* (1999), 61–85.

[17] D. Klatte and B. Kummer: Nonsmooth Equations in Optimization – Regularity, Calculus, Methods and Applications. Kluwer, Dordrecht – Boston – London 2002.

[18] D. Klatte and B. Kummer: Optimization methods and stability of inclusions in Banach spaces. Math. Program. Ser. B *117* (2009), 305–330.

[19] D. Klatte and K. Tammer: Strong stability of stationary solutions and Karush–Kuhn–Tucker points in nonlinear optimization. Ann. Oper. Res. *27* (1990), 285–307.

[20] M. Kojima: Strongly stable stationary solutions in nonlinear programs. In: Analysis and Computation of Fixed Points (S. M. Robinson, ed.), Academic Press, New York 1980, pp. 93–138.

[21] B. Kummer: Newton's method for non-differentiable functions. In: Advances in Math. Optimization (J. Guddat et al., eds.), Akademie Verlag, Berlin 1988, pp. 114–125.

[22] B. Kummer: Lipschitzian inverse functions, directional derivatives and application in $C^{1,1}$ optimization. J. Optim. Theory Appl. *70* (1991), 559–580.

[23] B. Kummer: An implicit function theorem for $C^{0,1}$-equations and parametric $C^{1,1}$-optimization. J. Math. Anal. Appl. *158* (1991), 35–46.

[24] B. Kummer: Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis. In: Advances in Optimization (W. Oettli and D. Pallaschke, eds.), Springer, Berlin 1992, pp. 171–194.

[25] B. Kummer: Generalized Newton and NCP-methods: Convergence, regularity, actions. Discuss. Math. – Differential Inclusions *20* (2000), 209–244.

[26] B. Kummer: Inclusions in Banach spaces: Hoelder stability, solution schemes and Ekeland's principle. J. Math. Anal. Appl. *358* (2009), 327–344.

[27] L. I. Minchenko: Multivalued analysis and differential properties of multivalued mappings and marginal functions. J. Math. Sci. *116* (2003), 93–138.

[28] R. A. Poliquin and R. T. Rockafellar: Tilt stability of a local minimum. SIAM J. Optim. *8* (1998), 287-299.

[29] S. M. Robinson: Strongly regular generalized equations. Math. Oper. Res. *5* (1980), 43–62.

[30] R. T. Rockafellar and R. J.-B. Wets: Variational Analysis. Springer, Berlin 1998.

[31] S. Scholtes: Introduction to Piecewise Differentiable Equations. Preprint No. 53/1994. Institut für Statistik und Math. Wirtschaftstheorie, Universität Karlsruhe, 1994.

[32] O. Stein: Bi-level Strategies in Semi-infinite Programming. Kluwer, Dordrecht – Boston – London 2003.

[33] K. Sydsaeter, P. Hammond, A. Seierstad, and A. Strom: Further Mathematics for Economic Analysis. Prentice Hall, 2005.

[34] L. Thibault: Subdifferentials of compactly Lipschitz vector-valued functions. Ann. Mat. Pura Appl. *4* (1980), 157–192.

[35] H. Varian: Microeconomic Analysis. Third edition. W. W. Norton, New York 1992.

Stephan Bütikofer, Institut für Datenanalyse und Prozessdesign, Zürcher Hochschule für Angewandte Wissenschaften, Rosenstrasse 3, CH-8400 Winterthur. Switzerland.
e-mail: stephan.buetikofer@zhaw.ch

Diethard Klatte, Institut für Operations Research, Universität Zürich, Moussonstrasse 15, CH-8044 Zürich. Switzerland.
e-mail: klatte@ior.unizh.ch

Bernd Kummer, Institut für Mathematik, Humboldt–Universität zu Berlin, Unter den Linden 6, D-10099 Berlin. Germany.
e-mail: kummer@mathematik.hu-berlin.de